The Role of Subclasses in Machine Diagnostics

Marina Skurichina, Alexander Ypma and Robert P.W.Duin

Pattern Recognition Group, Department of Applied Physics, Faculty of Applied Sciences,
Delft University of Technology, P.O. Box 5046, 2600GA Delft, The Netherlands.
marina@ph.tn.tudelft.nl

Abstract

In machine diagnostics it is difficult to collect for learning all possible operating modes of machine functioning. Some operating modes will usually be missing. In these circumstances, it is important to know which modes (subclasses) are the most valuable for successful machine diagnostics. It is also of interest to investigate the usefulness of noise injection to cover missing operating modes in the data. In this paper, we study the importance of selecting different operating modes of a water-pump and using them for learning in 2-class and 4-class problems. We show that the operating modes representing different running speeds are more valuable than those representing machine loads. We also demonstrate that 2-Nearest Neighbours directed noise injection is useful when filling in missing operating modes in the data.

1. Introduction

In machine diagnostics one is interested to get a full representation of all possible states of machine functioning: the normal behaviour - for a wide range of loads, speeds, environmental conditions and in time (for different ages of the machine) and the abnormal behaviour. However, in practice it is impossible to collect for learning all states of machine functioning. As a rule, collected observations (for discrete numbers of conditions, different ends of machine functioning, discrete life paths etc.) are sampled. Data are incomplete. Some operating modes (subclasses) are missing. Therefore, it becomes important to evaluate the significance of missing data for machine diagnostics. It is reasonable to assume that some subclasses may overlap. If missing subclasses overlap with the collected data, their absence will not alter much the quality of machine diagnostics. However, some missing subclasses may have unique information that is not presented in the collected data. Hence, their absence will affect machine diagnostics.

It is logical to assume that the data, describing machine functioning for gradually varying conditions (e.g., increasing or decreasing running speed or the machine load), will also slowly change in the data feature space, creating a sequence of subclasses. It might be of interest to study the cost of missing data in this sequence of subclasses for machine diagnostics, and also the possibility to regenerate missing data by noise injection. Other methods (clustering and time sequence) to recover lost information (but only contextual) have been considered by Turney [1]. Usually Gaussian noise is used in order to enrich available data. Here, Gaussian distributed objects are generated around each data object. However, when data samples are located in the subspace of the feature space where they are described, such Gaussian noise injection could deteriorate the real data distribution. In this case the k-Nearest Neighbours (k-NN) directed noise injection becomes more preferable [2] as it takes into account the distribution of available objects. By that, the new data objects are generated within borders of the distribution of existed objects, filling in missing links between observed objects.

In this paper we consider a number of situations, where different machine operating modes are available. We study the cost of the missing data in these situations for machine diagnostics. This study is carried out for 2-class and 4-class problems for the data [5] obtained from the water-pump. To evaluate the quality of machine diagnostics, the Parzen Window Classifier [3] and the Karhunen-Loeve Linear Classifier [4] are used. We also investigate the effectiveness of noise injection in regenerating missing subclasses in the data. In this study we compare two kinds of noise injection, 2-Nearest Neighbours (2-NN) directed noise injection (NI) and Gaussian noise injection (GNI).

2. Data

The data [5] consist of the measurements obtained from four kinds of water-pump operating states: a normal behaviour (NB), a bearing fault (BF) (a fault in the outer ring of the uppermost single bearing), an imbalance failure (IF) and a loose foundation failure (LFF), where the running

0-7695-0750-6/00 $10.00 © 2000 IEEE

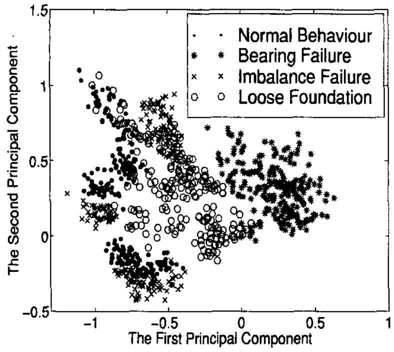

Figure 1. The scatter plot of the first two principal components of the autoregressive model coefficients for four operating states.

speed (46, 48, 50, 52 and 54 Hz) and machine load (25, 29 and 33 KW) are varied. To measure pump vibrations, a ring accelerometer is used. For the obtained time series of the pump vibration patterns, we determined the coefficients of an order 128 autoregressive model. The 128 coefficients of this model are used as the features describing the pump vibration patterns. For each operating mode 15 128-dimensional vectors are obtained, which are normalized w.r.t. the mean and the standard deviation. Then the data are combined either in 4 classes (a normal behaviour, a bearing fault, an imbalance failure and a loose foundation failure) consisted of 225 observations each, or in 2 classes (the normal behaviour and the abnormal behaviour). In the latter case, the normal behaviour class consisted of 225 observations the abnormal behaviour class consisted of 675 observations representing all three failures mentioned above.

The data dimensionality p=128 is large, compared to the number of available objects per subclass. However, the performed data analysis shows that the data are clustered and located in a subspace of the feature domain in which they are described (see Figure 1). Therefore, it is possible to reduce the data dimensionality. In order to perform it, we have applied the principal component analysis and decided to keep only 64 features with the largest variance.

3. Simulation Study

Let us study the importance of subclasses in machine diagnostics and the usefulness of noise injection in reconstructing missing subclasses in the data. We consider two kinds of noise injection (NI): Gaussian NI and 2-NN directed NI. The 2-NN directed NI is applied by us in the following way. At first, all available training objects are clustered by the k-means procedure inside each class C. Then, for the randomly chosen training object $X_j^{(i)}$ ($i=\overline{1,C}$; $j=\overline{1,N}$), two nearest objects to it, $Q_{j1}^{(i)}$ and $Q_{j2}^{(i)}$, are found, which belong to other cluster than $X_j^{(i)}$. Noise is generated only in positive directions of the nearest neighbours. It is defined as

$$Z_{jr}^{(i)} = \frac{1}{2}\sum_{k=1}^{2} \mathrm{sgn}(\xi_k)\xi_k\left(X_j^{(i)} - Q_{jk}^{(i)}\right) \quad (r=\overline{1,R})$$

where $\xi_k \sim U(\mathbf{0}, 1)$, R is the desired number of noise injections. We use R=100 per class for both kinds of NI.

In order to study the importance of subclasses in machine diagnostics, let us consider a number of cases (see Figure 2), where some subclasses (black ones) are selected for learning a classifier and other subclasses (crossed out) are used for testing. We intend to compare the quality of machine diagnostics on the complete data (the ideal case, when both black and crossed out subclasses are used for learning the classifier and testing), on incomplete data (the worst case, when black subclasses are used for learning, and crossed out subclasses are used for testing) and on incomplete data enriched by noise injection (when learning is performed on black subclasses enriched by noise injection, and testing is done on crossed out subclasses). In order to evaluate the quality of machine diagnostics, we use the Parzen Window Classifier (PWC) with the window width λ=0.1 and the Karhunen-Loeve Linear Classifier (KLLC), which constructs a linear discriminant in the subspace of the 8 features with the largest eigenvalues. For the ideal case, training sets are chosen randomly from all considered subclasses (black and crossed out) and the rest of the data is used as the test set. For all cases, experiments are repeated 25 times on independent randomly chosen training sets consisting of N objects per class. The mean classification errors obtained on the test set and their standard deviations (in brackets) are presented in tables 1-4. For 2-class problem posterior probabilities of classes are set

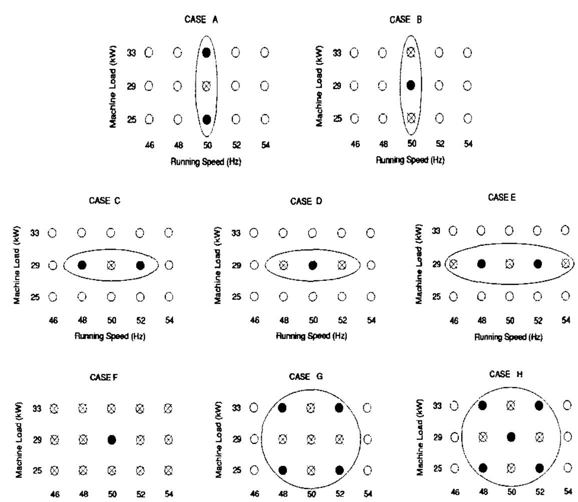

Figure 2. The incomplete data cases, where black subclasses are selected for learning, crossed out subclasses are used only for testing, white subclasses are not considered.

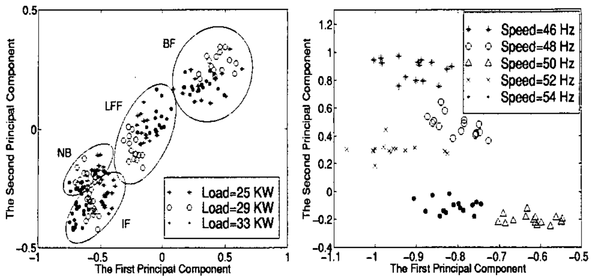

Figure 3. The scatter plot of different loads for four operating states with the running speed 50Hz and of different running speeds for normal behaviour with the machine load 29kW in the first two principal components domain.

0.25 for normal behaviour and 0.75 for abnormal behaviour.

The results of the simulation study performed by the PWC (tables 1 and 2) show that it is easier to get over missing machine loads (cases A, B) than over missing running speeds (cases C, D, E). It is happens because the subclasses representing different machine loads (for the same speed) are usually distributed nearby each other and partly overlapping (see Figure 3). Thus, interpolation and even extrapolation of the loads is not a problem. However, the subclasses representing different running speeds (for the same load) may sometimes be away from each other. They also may be distributed in the feature space inconsistently to the values of the running speed (see Figure 3). By this reason, even interpolation of running speeds seems to be a real problem. One can see, that 2-NN directed NI can improve the situation and sometimes solve the problem (e.g., cases C, G for 4-class problem). On the example of cases G and H, we see that having only the bound subclasses is not enough to interpolate missing data. In case H, compared to case G, we use an additional subclass for learning. It reasonably improves machine diagnostics. As enough different loads are presented in the data, one can interpolate and extrapolate the missing subclasses well. Case F shows an interesting result, which demonstrates that even one properly chosen subclass (typical running speed and load) allows us extrapolate successfully a large number of missing subclasses. Concerning two different types of noise injection, one can see that, for the PWC, 2-NN NI was very helpful in regenerating missing subclasses, just as Gaussian NI was completely useless We should also point that the performance of the PWC was some what worse than for 4-class problem. It happens because classes representing abnormal situations are not much overlapping, just as the class of the normal behaviour has overlaps with other classes. Therefore, combining abnormal situations in one class has worsened machine diagnostics.

Comparing the performance of the PWC (tables 1 and 2) and the KLLC (tables 3 and 4), one can see that sometimes the KLLC performs worse for the ideal situation. However, it gives relatively better results in the worst situation. Noise injection is not very useful for the KLLC.

Table 1. The performance of the Parzen Window Classifier for 4-class problem.

case	N	ideal	worst	2-NN NI	GNI
A	10	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
A	25	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
B	10	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
C	10	0.033 (0.013)	0.271 (0.009)	0.087 (0.014)	0.273 (0.01)
C	25	0.0 (0.0)	0.263 (0.002)	0.034 (0.007)	0.261 (0.002)
D	10	0.021 (0.008)	0.129 (0.002)	0.138 (0.004)	0.129 (0.002)
E	10	0.076 (0.007)	0.303 (0.005)	0.252 (0.006)	0.306 (0.006)
E	25	0.003 (0.003)	0.30 (0.002)	0.242 (0.003)	0.303 (0.003)
F	10	0.091 (0.008)	0.182 (0.001)	0.177 (0.001)	0.182 (0.001)
G	10	0.0185 (0.004)	0.168 (0.004)	0.098 (0.007)	0.167 (0.004)
G	25	0.0015 (0.000)	0.178 (0.001)	0.089 (0.005)	0.176 (0.002)
H	10	0.0135 (0.002)	0.036 (0.011)	0.044 (0.01)	0.037 (0.011)
H	25	0.0014 (0.001)	0.002 (0.001)	0.008 (0.002)	0.003 (0.001)

Table 2. The performance of the Parzen Window Classifier for 2-class problem.

case	N	ideal	worst	2-NN NI	GNI
A	10	0.001 (0.001)	0.004 (0.003)	0.005 (0.004)	0.004 (0.002)
A	25	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
B	10	0.0 (0.0)	0.0 (0.0)	0.01 (0.0)	0.0 (0.0)
C	10	0.107 (0.011)	0.22 (0.013)	0.143 (0.017)	0.221 (0.013)
C	25	0.016 (0.007)	0.235 (0.004)	0.179 (0.016)	0.235 (0.005)
D	10	0.093 (0.013)	0.112 (0.012)	0.06 (0.01)	0.112 (0.012)
E	10	0.182 (0.01)	0.241 (0.009)	0.168 (0.012)	0.24 (0.009)
E	25	0.073 (0.011)	0.244 (0.005)	0.163 (0.009)	0.245 (0.004)
F	10	0.166 (0.012)	0.141 (0.008)	0.085 (0.004)	0.141 (0.008)
G	10	0.09 (0.012)	0.145 (0.008)	0.12 (0.009)	0.145 (0.008)

case	N	ideal	worst	2-NN NI	GNI
G	**25**	0.022 (0.006)	0.118 (0.002)	0.106 (0.007)	0.118 (0.003)
H	**10**	0.107 (0.014)	0.093 (0.012)	0.091 (0.012)	0.094 (0.012)
H	**25**	0.009 (0.004)	0.034 (0.01)	0.044 (0.01)	0.033 (0.01)

Table 3. The performance of the Karhunen-Loeve Linear Classifier for 4-class problem.

case	N	ideal	worst	2-NN NI	GNI
A	**10**	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
A	**25**	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
B	**10**	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
C	**10**	0.033 (0.07)	0.191 (0.022)	0.092 (0.02)	0.205 (0.024)
C	**25**	0.004 (0.001)	0.101 (0.01)	0.058 (0.015)	0.122 (0.014)
D	**10**	0.034 (0.007)	0.139 (0.006)	0.138 (0.007)	0.142 (0.008)
E	**10**	0.15 (0.01)	0.277 (0.008)	0.228 (0.009)	0.284 (0.01)
E	**25**	0.109 (0.008)	0.264 (0.005)	0.212 (0.008)	0.265 (0.007)
F	**10**	0.166 (0.007)	0.194 (0.003)	0.194 (0.003)	0.197 (0.004)
G	**10**	0.046 (0.005)	0.109 (0.007)	0.086 (0.007)	0.113 (0.007)
G	**25**	0.025 (0.001)	0.096 (0.007)	0.087 (0.006)	0.104 (0.008)
H	**10**	0.047 (0.004)	0.065 (0.006)	0.057 (0.01)	0.069 (0.007)
H	**25**	0.025 (0.002)	0.041 (0.003)	0.031 (0.003)	0.042 (0.003)

Table 4. The performance of the Karhunen-Loeve Linear Classifier for 2-class problem.

case	N	ideal	worst	2-NN NI	GNI
A	**10**	0.002 (0.001)	0.003 (0.002)	0.003 (0.002)	0.001 (0.001)
A	**25**	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
B	**10**	0.005 (0.003)	0.0 03 (0.001)	0.003 (0.002)	0.003 (0.001)
C	**10**	0.096 (0.011)	0.111 (0.018)	0.101 (0.019)	0.109 (0.02)
C	**25**	0.04 (0.009)	0.079 (0.015)	0.128 (0.016)	0.095 (0.015)
D	**10**	0.07 (0.011)	0.105 (0.012)	0.11 (0.017)	0.115 (0.016)
E	**10**	0.181 (0.013)	0.164 (0.01)	0.151 (0.01)	0.163 (0.012)
E	**25**	0.157 (0.012)	0.149 (0.009)	0.15 (0.011)	0.164 (0.01)
F	**10**	0.171 (0.009)	0.153 (0.006)	0.157 (0.008)	0.154 (0.007)
G	**10**	0.091 (0.012)	0.109 (0.012)	0.136 (0.01)	0.115 (0.012)
G	**25**	0.046 (0.006)	0.078 (0.009)	0.096 (0.013)	0.088 (0.009)
H	**10**	0.085 (0.012)	0.079 (0.01)	0.102 (0.012)	0.083 (0.01)
H	**25**	0.045 (0.006)	0.064 (0.009)	0.063 (0.009)	0.066 (0.009)

4. Conclusions

When performing machine diagnostics, it is impossible to have all possible operating modes for learning. Some operating modes usually are missing. Therefore, it is important to understand which operating modes are important to have in the training set for successful machine diagnostics, and which operating modes can be reconstructed from the available modes by noise injection. The performed simulation study on the coefficients of the autoregressive model constructed from the water-pump vibration patterns has shown that running speeds are much more important in a water-pump diagnostics than machine loads. Consequently, to regenerate missing operating modes between machine loads is easier than between running speeds. In order to reconstruct missing subclasses in the data, 2-NN directed noise injection can be used. It can reasonably improve the performance of the PWC, when it is used for machine diagnostics.

Acknowledgment

This work is supported by the Foundation for Applied Sciences (STW) and the Dutch Organization for Scientific Research (NWO).

References

1. P. Turney. The management of context-sensitive features: a review of strategies. *Proceedings of the Workshop on Learning in Context-Sensitive Domains (at the ICML-96)*: 60-66, Bari, Italy, 1996.
2. M. Skurichina, S. Raudys and R.P.W. Duin. K-nearest neighbours noise injection in multilayer perceptron training. *IEEE Transaction on Neural Networks*, 11 (2): 504-511, 2000.
3. E. Parzen. On estimation of a probability density function and mode. *Annals of Math.Statistics*, 33: 1065-1076, 1962.
4. K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, San-Diego, 1990.
5. A.Ypma, R. Ligteringen, E.E.E. Frietman and R.P.W. Duin. Recognition of bearing failures using wavelets and neural networks. *Proceedings of the TFT'97*, Coventry, UK, 1997.